Analysis of Lempel-Ziv'78 for Markov Sources

Philippe Jacquet, Wojciech Szpankowski

Analysis of Lempel-Ziv'78 for Markov Sources

Philippe Jacquet
INRIA, Paris, France
philippe.jacquet@inria.fr

Wojciech Szpankowski
Center for Science of Information, Department of Computer Science,
Purdue University, West Lafayette, IN, USA
spa@cs.purdue.edu

Abstract

Lempel-Ziv'78 is one of the most popular data compression algorithms. Over the last few decades fascinating properties of LZ78 were uncovered. Among others, in 1995 we settled the Ziv conjecture by proving that for a *memoryless source* the number of LZ78 phrases satisfies the Central Limit Theorem (CLT). Since then the quest commenced to extend it to Markov sources. However, despite several attempts this problem is still open. The 1995 proof of the Ziv conjecture was based on two models: In the DST-model, the associated digital search tree (DST) is built over m *independent* strings. In the LZ-model a *single* string of length n is partitioned into variable length phrases such that the next phrase is not seen in the past as a phrase. The Ziv conjecture for memoryless source was settled by proving that both DST-model and the LZ-model are asymptotically equivalent. The main result of this paper shows that this is not the case for the LZ78 algorithm over Markov sources. In addition, we develop here a large deviation for the number of phrases in the LZ78 and give a *precise* asymptotic expression for the redundancy which is the excess of LZ78 code over the entropy of the source. We establish these findings using a combination of combinatorial and analytic tools. In particular, to handle the strong dependency between Markov phrases, we introduce and precisely analyze the so called *tail symbol* which is the first symbol of the next phrase in the LZ78 parsing.

2012 ACM Subject Classification Mathematics of computing → Information theory

Keywords and phrases Lempel-Ziv algorithm, digital search trees, depoissonization, analytic combinatorics, large deviations

Digital Object Identifier 10.4230/LIPIcs.AofA.2020.15

Funding *Wojciech Szpankowski*: This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grant CCF-1524312.

Acknowledgements We thank Guillaume Duboc for simulation of LZ78 scheme resulting in Figure 2.

1 Introduction

The Lempel-Ziv compression algorithm [16] is a universal compression scheme. It partitions the text to be compressed into consecutive phrases such that the next phrase is the unique shortest prefix (of the uncompressed text) not seen before as a phrase. For example, $aababbababbb$ is parsed as $()(a)(ab)(abb)(aba)(b)(bb)$. The LZ78 compression code consists of a pointer to the previous phrase and the last symbol of the current phrase. The distribution of the number of phrases and other related quantities (such as redundancy and code length) are known for memoryless sources [10, 14] but research over the past 40 years has failed to produce any significant progress for Markov sources. In this paper, we present novel large deviations and precise redundancy results that had been wanting since the algorithm inception, as well as some surprising findings regarding the difference between the memoryless case and the Markov case.

It is convenient to organize phrases (dictionary) of the Lempel-Ziv scheme in a *digital search tree* (DST) [6] which represents a parsing tree. We assume throughout that $\mathcal{A} = \{a, b\}$. Then the root contains an empty phrase. The first phrase is the first symbol, say "$a \in \mathcal{A}$" which is stored in a node appended to the root. The next phrase is either $(aa) \in \mathcal{A}^2$ stored in another node that branches out from the node containing the first phrase "a" or (ab) that is stored in a node attached to the root. This process repeats recursively until the text is parsed into full phrases (see Figure 1). A detailed description can be found in [3, 6, 8].

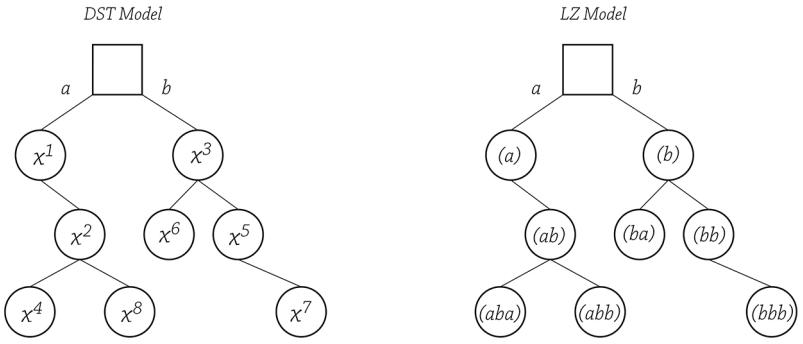

Figure 1 The DST-model vs LZ-model. In the DST-model we inserted eight (infinite) strings: $X^1 = a\mathbf{b}b\cdots$, $X^2 = ab\mathbf{b}\cdots$, $X^3 = bb\mathbf{b}a\cdots$, $X^4 = aba\mathbf{a}a\cdots$, $X^5 = bb\mathbf{a}a\cdots$, $X^6 = ba\mathbf{a}a\cdots$, $X^7 = bbb\mathbf{a}\cdots$ and $X^8 = abb\mathbf{b}b\cdots$, where bold symbols denote DST tail symbols. In the LZ-model we parsed one string $X = ()(a)(\mathbf{a}b)(\mathbf{b})(\mathbf{a}aba)(\mathbf{b}b)(\mathbf{b}bb)(\mathbf{a}bb)$ with bold denoting LZ tail symbols.

We consider two models called the DST-model and the LZ-model. In the DST-model we insert *independent strings* although each string may be generated by a source with memory like a Markov source. In the LZ-model we parse a *single* string as shown in Figure 1. We distinguish two types of DST and LZ models. To define them we need to introduce the path length L as the sum of all depths in the digital search tree or the sum of all phrases in the LZ model. In the "m"-DST model we insert m independent strings into a digital search tree – leading to a variable path length denoted as L_m – while the "n"-DST model is built over a random number of independent strings such that the total path length is equal to n. Similarly, we have "m"-LZ and "n"-LZ models: In the former we construct m LZ phrases to form a string of (variable) length denoted as $\mathcal{L}_m$ while in the "n"-LZ model we parse a string of length n into a variable number of phrases that we denote as M_n. Throughout, m will denote number of strings or phrases while n will stand for the length of a string.

There is a simple relation between M_n and $\mathcal{L}_m$ called the *renewal equation* which asserts

$$P(M_n > m) = P(\mathcal{L}_m < n). \tag{1}$$

Finally, observe that the code length of the LZ78 algorithm is $C_n = \sum_{k=1}^{M_n} \lceil \log_2(k) \rceil + \lceil \log_2(|\mathcal{A}|) \rceil$ since the pointer to the kth node requires at most $\lceil \log_2 k \rceil$ bits, while the next symbol costs $\lceil \log_2 |\mathcal{A}| \rceil$ bits. For binary alphabet $\mathcal{A} = \{a, b\}$ we simplify the code length to $C_n = M_n (\log_2 M_n + 1)$.

To understand LZ78 behavior one must analyze the limiting distribution of M_n and/or $\mathcal{L}_m$ connected through the renewal equation (1). For *memoryless* sources we benefited from the fact the random variable L_m and $\mathcal{L}_m$ are *probabilistically equivalent* as shown in 1995 paper [3]. Unfortunately, this equivalence breaks for sources with memory such as Markov sources. To capture this dependency we introduce the notion of the *tail symbol*. In the

DST-model the tail symbol of an inserted string is the first non-inserted symbol of that string, as shown in Figure 1. In the LZ-model the tail symbol of a phrase is the first symbol of the next phrase (see Figure 1). Furthermore, in the Markov case there is additional complication, even for the DST-model. In the DST-model we need to consider two digital search trees: one built over all (independent) strings starting with symbol $a \in \mathcal{A}$, and the second one built over all strings that start with $b \in \mathcal{A}$. At the end we construct a cumulative knowledge by weighting over the initial symbols (see [7]).

In this paper, we present large deviation results for the number of phrases M_n in "n"-LZ model and the average length of a LZ (Markov) string built over m phrases in the "m"-LZ model.[1] In the memoryless case we could read the number of phrases M_n directly from the path length L_m of the m-DST model. It is *not* the case in the Markov model but through the tail symbol distribution we will connect both quantities. Recall that $\mathcal{L}_m$ is the length of a string generated by a Markov source which is parsed by the LZ78 scheme until we see m phrases (our m-LZ model). This should be compared to the total path length L_m (notice roman font for L) in the the m-DST model. In the memoryless case, we proved in [3, 5] that the expected value of L_m and the expected value of the length of a string built from m phrases, $\mathcal{L}_m$, are the same. Somewhat surprisingly it is not the case for the Markov case. We will prove in Theorem 5 that $\mathbf{E}[L_m] - \mathbf{E}[\mathcal{L}_m] = \Theta(m)$.

Let us now briefly review literature on LZ78 and DST analysis. The goal is to prove the Central Limit Theorem (CLT) for the number of phrases and establish precise rate of decay of the LZ78 code redundancy for Markov sources. For memoryless sources, CLT was already proved in [3] while the average redundancy was presented in [10, 14]. It should be pointed out that since 1995 paper [3] no simpler, in fact, no new proof of CLT was presented except the one by Neininger and Rüschendorf [13] but only for *unbiased* memoryless sources (as in [1]). The only known to us analysis of LZ78 for Markov sources is presented in [7], but the authors restricted their attention to a single phrase. We should point out that for another Lempel-Ziv scheme known as LZ'77 algorithm, Fayolle and Ward [2] analyzed an associated suffix tree built over a Markov string and obtained the distribution of the depth, which allows us to conclude the limiting distribution of a phrase in the LZ'77 scheme (see also [11, 12]). Regarding analysis of digital search trees, and in general digital trees, more is known [8, 6, 15]. Digital trees for memoryless sources were analyzed in [1, 10, 6] while digital trees under Markovian models were studied in [7, 9, 2]. This information is surveyed in detail in [6].

The paper is organized as follows. In the next section we present our main results regarding the LZ and DST models including the mean, variance and distribution of the number of tail symbols in the DST model (see Theorem 2–4), and large deviations as well as precise redundancy for the LZ model (see Theorems 5–6). We prove these findings in Section 3 (DST model) and in Section 4 (LZ model), with most details delayed till the appendix. Throughout we use combinatorics on words and analytic tools such as generating functions, Poisson transform, analytic depoissonization, and Mellin transform.

2 Main Results

We consider a stationary ergodic Markov source generating a sequence of symbols drawn from a finite alphabet $\mathcal{A}$. In this paper we study only a binary Markovian process of order 1 with the transition matrix $\mathbf{P} = [P(c|d)]_{c,d\in\mathcal{A}}$ where $\mathcal{A} = \{a, b\}$. In this section we present our main results with proof delayed till Sections 3–4 and appendix. However, first we present a road map of our methodology and findings.

[1] From now on we drop the quotes around m and n to simplify the presentation.

Our main goal is to analyze the Lempel-Ziv'78 scheme for Markovian input. However, as discussed before, we first consider an auxiliary model named DST-model built over m independent Markov strings, also called the m-DST model. However, for Markov sources we need to construct two *conditional* digital search trees: one built over m Markov strings all starting with symbol $a \in \mathcal{A}$ and the other DST built over m strings starting with $b \in \mathcal{A}$. We write $c \in \mathcal{A}$ for a generic symbol from $\mathcal{A}$, that is, either $c = a$ or $c = b$. For a given $c \in \mathcal{A}$, we consider m independent Markov strings all starting with c and build an m-DST tree. For such a tree we analyze two quantities, namely the total path length denoted as L_m^c, and the number $T_m^c(a)$ of inserted strings (all starting with c) with the tail symbol a, that is, among m Markov strings there are $T_m^c(a)$ strings with the tail symbol a. Clearly, $T_m^c(a) + T_m^c(b) = m$. Throughout, we also assume that the tail symbol is always a so we just write $T_m^c := T_m^c(a)$. In Theorems 2-3 we summarize our new results regarding T_m^c, while in Theorem 4 we present large deviation results for both T_m^c and L_m^c.

Second, we consider the m-LZ model (in which we run LZ78 algorithm on a single string until we see m phrases) and tie it up to the m-DST model just discussed. Here we use a combinatorial approach. For a given sequence $\mathbf{s}$ over $\mathcal{A}$ of length m we compare in Lemmas 10-11 two probabilities: (i) the probability that in the m-LZ model (constructed from m LZ phrases) we end up with a LZ sequence of length n having all tail symbols equal to $\mathbf{s}$; and (ii) the probability that in the m-DST model (built over m independent Markov strings) the resulting digital search tree has path length equal to n and all tail symbols are equal to $\mathbf{s}$. Using this, we present in Theorem 5 our large deviations for the m-LZ model and using the renewal equation (1) in Theorem 6 we establish large deviations for the n-LZ model. In Corollary 7 we find a *precise* expression for the redundancy of LZ78 for Markov sources.

Finally, when comparing the average path length L_m^c in the m-DST model with the length $\mathcal{L}_m^c$ in the m-LZ model we shall use the following simple fact.

▶ **Proposition 1.** *For $\delta < 1$ let there exist $B, C > 0$ such that for a discrete random variable X_m the following holds uniformly*

$$P(X_m = k) \le B \exp\left(-Cm^{-\delta}|k - A_m|\right). \tag{2}$$

Then

$$\boldsymbol{E}[X_m] = A_m + O(m^\delta). \tag{3}$$

Proof. Define $B_m = m^\delta (\log B)/C \le |k - A_m|$. Then it is easy to see that $EX_m = \sum_k kP(X_m = k) = A_m + \sum_k (k - A_m)P(X_m = k)$, and the latter term can be estimated by the integral $2B \int_0^\infty \exp(-Cm^{-\delta}x)(x+1)dx = O(m^\delta)$. This completes the proof. ◀

2.1 Results on DST

In this section we summarize our results for the m-DST model: We first focus on the number of times, $T_m^c := T_m^c(a)$, the tail symbol is a when all m Markov sequences start with $c \in \mathcal{A}$. Then we study the path length L_m^c in the m-DST model when all sequences start with c. Finally, we present large deviations for both T_m^c and L_m^c.

For $c \in \mathcal{A}$, let $D_m^c(u) = E[u^{T_m^c}]$ be the probability generating function of T_m^c defined for a complex variable u. We have the recursion:

$$D_{m+1}^c(u) = (P(a|c)u + 1 - P(a|c)) \sum_k \binom{m}{k} P(a|c)^k P(b|c)^{m-k} D_k^a(u) D_{m-k}^b(u) \tag{4}$$

subject to $D_0^c(u) = 1$ and $D_1^c(u) = P(a|c)u + 1 - P(a|c)$. Furthermore, define the bivariate Poisson transform $D_c(z,u) = \sum_{m\geq 0} \mathbf{E}[u^{T_m^c}]\frac{z^m}{m!}e^{-z}$. From above we easily find the following differential-functional equation

$$\partial_z D_c(z,u) + D_c(z,u) = D_1^c(u) D_a(P(a|c)z, u)) \cdot D_b(P(b|c)z, u) \tag{5}$$

with $D_c(z,1) = 1$ where ∂_z is the partial derivative with respect to variable z.

We now focus on the first Poisson moment $X_c(z) = \partial_u D_c(z,1)$ where ∂_u is the derivative with respect to variable u. We also study the Poisson variance $V_c(z) = \partial_u^2 D_c(z,1) + X_c(z) - (X_c(z))^2$, and the limiting distribution of T_m^c. After finding the asymptotic behavior of the Poisson mean $X_c(z)$ and variance $V_c(z)$ for large $z \to \infty$ we invoke the depoissonization lemma of [4] to extract the original mean and variance:

$$\mathbf{E}[T_m^c] = X_c(m) - \frac{1}{2} m \partial_z X_c(m) + O(X_c(m)/m), \quad \mathrm{Var}[T_m^c] \sim V_c(m) - m[\partial_z X_c(m)]^2.$$

Let us start with the Poisson mean $X_c(z)$. Taking the derivative of (5) with respect to u and setting $u = 1$ we find

$$\partial_z X_c(z) + X_c(z) = P(a|c) + X_a(P(a|c)z) + X_b(P(b|c)z). \tag{6}$$

To complete this equation we need to calculate the initial values of $\mathbf{E}[T_m^c]$. It is easy to see that

$$\mathbf{E}[T_0^c] = 0, \quad \mathbf{E}[T_1^c] = P(a|c), \quad \mathbf{E}[T_2^c] = P(a|c) + P(a|c)P(a|a) + P(b|c)P(a|b). \tag{7}$$

In a similar fashion we can derive the differential-functional equation for the Poisson variance. After some tedious algebra we arrive at

$$\partial_z V_c(z) + V_c(z) = P(a|c) - P^2(a|c) + [\partial_z X_c(z)]^2 + V_a(P(a|c)z) + V_b(P(b|c)z). \tag{8}$$

Both differential-functional system of equations (5) and (7) can be solved using complicated Mellin transform approach [15]. We provide details of our approach in the Appendix. For now we need to introduce some extra notation to present our main results. For complex s define

$$\mathbf{P}(s) = \begin{bmatrix} P(a|a)^{-s} & P(b|a)^{-s} \\ P(a|b)^{-s} & P(b|b)^{-s} \end{bmatrix}. \tag{9}$$

For such $\mathbf{P}(s)$ we denote by $\lambda(s)$ the main eigenvalue and $\boldsymbol{\pi}(s)$ the main eigenvector. We notice that $\boldsymbol{\pi}(-1)$ is the stationary vector of the Markov process. We also need another matrix

$$\mathbf{Q}(s) = \prod_{i\geq 1} (\mathbf{I} - \mathbf{P}(s-i))^{-1} \prod_{j=-\infty}^{j=-2} (\mathbf{I} - \mathbf{P}(j))$$

defined for $\Re(s) \in (-2, 0)$. Furthermore, $\langle \mathbf{x}, \mathbf{y} \rangle$ is the scalar product of vectors $\mathbf{x}$ and $\mathbf{y}$.

Now we are in the position to formulate our main result.

▶ **Theorem 2.** *Consider a digital search tree built over m independent sequences (m-DST) generated by a Markov source. We have $\boldsymbol{E}[T_m^c] = \tau_c(m)m$ and $\boldsymbol{E}[L_m^c] = m\log m/h + m + \mu_c(m)m$ such that:*

- $\tau_c(m+1) - \tau_c(m) = O(1/m)$ *and* $\mu_c(m+1) - \mu_c(m) = O(1/m)$
- $\forall (c,d) \in \mathcal{A}^2$ $\tau_c(m) - \tau_d(m) = O(1/m)$ *and* $\mu_c(m) - \mu_d(m) = O(1/m)$.

Thus $\tau_c(m) = \tau(m) + O(1/m)$ where $\tau(m)$ does not depend on initial symbol c. In fact, $\tau(m)$ depends on the tail symbol, but since throughout the paper we assume the tail symbol is always a, we drop this dependency on a in $\tau(m)$. We present precise formula on $\tau(m)$ in the next theorem.

Similarly we have $\mu_c(m) = \mu(m) + O(1/m)$. The function $\mu(m)$ for Markov sources is given in Theorem 1 of [7]. For the memoryless source, it is $\frac{h_2}{h} + \gamma - 1 + \alpha$ and the average path length is $m \log m/h + m\mu(m)$, as discussed in [3].

To complete our analysis of the tail symbol, we present now precise behaviour of $\tau(m)$. We give a detailed proof in the Appendix.

▶ **Theorem 3.** *For $(a, b, c) \in \mathcal{A}^3$ define*

$$\alpha_{abc} = \log\left[\frac{P(a|b)P(c|a)}{P(c|b)}\right]. \tag{10}$$

(i) Aperiodic case. *If not all $\{\alpha_{abc}\}$ are rational, then $\tau(m) = \bar{\tau} + o(1)$ with*

$$\bar{\tau} = \pi_a + \frac{1}{\lambda'(-1)}\langle\left(\boldsymbol{\pi}'(-1) + \boldsymbol{\pi}\boldsymbol{Q}'(-1)\right)(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{P}\boldsymbol{e}_a\rangle, \tag{11}$$

where π_a is the stationary distribution of symbol a, and $\boldsymbol{e}_a$ is the vector made of a single 1 at the position corresponding to symbol a and zero otherwise.

Periodic case. *If all $\{\alpha_{abc}\}$ are rationally related, then for some $\varepsilon > 0$ we have $\tau(m) = \bar{\tau}(m) + O(m^{-\varepsilon})$ with $\bar{\tau}(m) = \bar{\tau} + Q_1(\log m)$, where $Q_1(.)$ is a periodic function.*

(ii) Variance. *The variance $\mathrm{Var}[T_m^c]$ grows linearly, that is $\mathrm{Var}[T_m^c] \sim m\omega_a(m)$, where $\omega_a(m) = \bar{\omega}_a$ for the aperiodic case and $\omega_a(m) = \bar{\omega}_a + Q_2(m)$ for the periodic case, where $\bar{\omega}_a$ is given explicitly in the Appendix in (B.16) of Theorem 14, and $Q_2(m)$ is a nonzero periodic function for rationally related case, and zero otherwise.*

(iii) Central Limit Theorem. *For any $c \in \mathcal{A}$ we have*

$$\frac{T_m^c - \boldsymbol{E}[T_m^c]}{\mathrm{Var}[T_m^c]} \xrightarrow{d} N(0, 1)$$

where $N(0, 1)$ denotes the standard normal distribution.

Similarly we have the same behaviour for $\mu(m)$ which is equal to $\bar{\mu} + o(1)$ in the aperiodic case and, in the periodic case, is equal to $\bar{\mu} + Q_3(\log m) + O(m^{-\varepsilon})$ whose expressions are in [3] and [7] where $Q_3(.)$ is a periodic function. For details the reader is referred to [7].

We notice that, unexpectedly, the number of tail symbols equal to a *is not* converging to $n\pi_a$ as we should expect from a Markovian sequence. The reason is that the tail symbol is not picked up at random in the sequence but occurs when the sequence path leaves the tree.

Finally, we present joint large deviations for both T_m^c and L_m^c which is a new result needed to establish large deviations for the LZ model. We prove it in Section 3.

▶ **Theorem 4.** *Consider a digital search tree (DST) built over m independent sequences generated by a Markov source. For all $\delta > 1/2$ there exist B, C and β strictly positive such that for all $x > 0$ uniformly in x*

$$P\left(|T_m^c - \boldsymbol{E}[T_m^c]| + |L_m^c - \boldsymbol{E}[L_m^c]| \geq xm^\delta\right) \leq Be^{-xCm^\beta} \tag{12}$$

for large m.

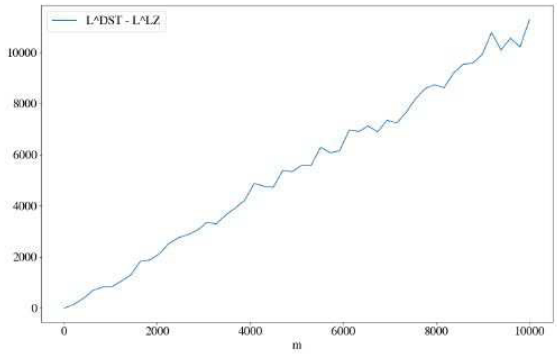

Figure 2 The difference $\mathbf{E}[L_m^c] - \mathbf{E}[\mathcal{L}_m^c]$ by simulation confirming that it grows linearly with m.

2.2 Results for the LZ78 Model

Let us start with the m-LZ model. For a given m, let $\mathcal{L}_m^c$ (note calligraphic $\mathcal{L}$) be the length of the LZ78 string composed of m phrases when the first phrase starts with symbol c. For memoryless sources, this quantity is equivalent to the path length L_m in the associated DST built over m independent strings. However, it is not the case for Markov sources. In Section 4 we prove Theorem 5 presented below by showing that $\mathbf{E}[L_m^c] - \mathbf{E}[\mathcal{L}_m^c] = \Theta(m)$, unlike in the memoryless case. Figure 2 compares the difference $\mathbf{E}[L_m^c] - \mathbf{E}[\mathcal{L}_m^c]$ obtained by simulation results confirming our theoretical findings.

▶ **Theorem 5.** *For m given, let $m^* := m^*(m)$ be the root of $x - x\tau(x) - (m-x)\tau(m-x)$.*

(i) *The average length $\boldsymbol{E}[\mathcal{L}_m^c]$ of the LZ-sequence consisting of the first m phrases is (for the aperiodic case)*

$$\boldsymbol{E}[\mathcal{L}_m^c] = m\log m/h + \mu(m^*)m^* + \mu(m-m^*)(m-m^*) + m(1-H(m^*/m)/h) + O(m^\delta) \tag{13}$$

where $H(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy and h the source entropy.

(ii) *For all $\delta > 1/2$ there exist $B, C, \beta > 0$, and $\gamma > 0$ such that uniformly for all $x > 0$*

$$P\left(|\mathcal{L}_m^c - E[\mathcal{L}_m^c]| \geq xm^\delta\right) \leq Bm^\gamma e^{-xCm^\beta} \tag{14}$$

for large m.

▶ Remark. The property of function $\tau(\cdot)$ implies that the equation $x - x\tau(x) - (m-x)\tau(m-x)$ has a single root as we will see in the proof of Section 4. Notice that m^*/m converges to $\bar{\tau}$ in the aperiodic case, and similarly $\mu(m^*)m^* + \mu(m-m^*)(m-m^*)$ is asymptotically equivalent to $\bar{\mu}m$. In the periodic case there will be small periodic contributions (contained in $\tau(m)$ and $\mu(m)$) as shown in Theorem 3. Notice that $H(m^*/m)$ is the tail symbol entropy, which is equal to h when the source is memoryless.

Our next goal is to present large deviation for the number of LZ phrases in the n-LZ model. Let M_n^c be the number of phrases obtained by parsing a Markovian sequence of length n starting with symbol c. By the renewal equation (1) we have $P(M_n^c > m) = P(\mathcal{L}_m^c < n)$ for all legitimate m and n. This allows us to read large deviation of M_n^c from Theorem 5. Following the footsteps of Theorem 2 of [5] we arrive at our next main result.

▶ **Theorem 6.** *For all $\delta > 1/2$ there exist B, C, β, and γ all strictly positive such that*

$$P\left(|M_n^c - \ell_c^{-1}(n)| \geq xn^\delta\right) \leq Bn^\gamma e^{-xCn^\beta}$$

where $\ell_c^{-1}(.)$ is the inverse function of $\ell_c(m) = \ell(m) + o(1)$ defined as $\ell(m) = \frac{m}{h}(\log m + \beta(m))$ with

$$\beta(m) = h\mu(m^*)m^*/m + h\mu(m - m^*)(m - m^*)/m - h + H(m^*/m)$$

where m^* is defined in Theorem 5 and $\mu(m)$ has extra fluctuating function in the periodic case.

Using Theorem 6 we can find a precise estimate on the LZ78 redundancy. Indeed, a good approximation for the LZ78 code length is $C_n^c = M_n^c(\log M_n^c + 1)$. The average conditional redundancy is defined as $r_n^c := \mathbf{E}[C_n^c]/n - h$, while the total average redundancy is $r_n = \pi_a r_n^a + \pi_b r_n^b$.

▶ **Corollary 7.** *The average redundancy rate r_n satisfies for all $\frac{1}{2} < \delta < 1$:*

$$r_n = h\frac{1 - \beta(\ell^{-1}(n))}{\log \ell^{-1}(n) + \beta(\ell^{-1}(n))} + O(n^{\delta-1}\log n) \sim h\frac{1 - \beta(\ell^{-1}(n))}{\log n},$$

and more specifically in the aperiodic case we have

$$r_n \sim h\frac{1 - \bar{\mu}}{\log n} + \frac{H(\bar{\tau}) - h}{\log n}$$

where $m^*/m \to \bar{\tau}$.

3 Proof of Theorem 4 for DST

Now we prove Theorem 4, that is, the joint large deviations for T_m^c and L_m^c in the m-DST model. We use Chernoff's bounds, so we need to introduce some bivariate generating functions. Define $P_{m,k,\ell}^c = P(T_m^c = k \ \& \ L_m^c = \ell)$, $P_m^c(u,v) = \mathbf{E}[u^{T_m^c} v^{L_m^c}] = \sum_{k,\ell} P_{m,k,\ell}^c u^k v^\ell$ and $P_c(z,u,v)$ to be the Poisson generating function $P_c(z,u,v) = \sum_m P_n^c(u,v)\frac{z^m}{m!}e^{-z}$. The following partial differential equation for $P_c(z,u,v)$ is easy to establish from (5)

$$\partial_z P_c(z,u,v) + P_c(z,u,v) = (uP(a|c) + P(b|c))P_a(P(a|c)zv,u,v)P_b(P(b|c)zv,u,v).$$

Lemma below is equivalent to Theorem 10 of [5] so we skip the proof in this conference paper.

▶ **Lemma 8.** *For all reals $\varepsilon' > 0$ and $\epsilon > 0$, there exists $0 < \vartheta < \pi/2$ and a complex neighborhood $\mathcal{U}(0)$ of 0 such that iuniformly for $(t_1,t_2) \in \mathcal{U}(0)^2$ and $|\arg(z)| < \vartheta$ so that $\log(P_c(z, e^{t_1|z|^{-\varepsilon'}}, e^{t_2|z|^{-\varepsilon'}}))$ exists and $\log(P_c(z, e^{t_1|z|^{-\varepsilon'}}, e^{t_2|z|^{-\varepsilon'}}) = O(z^{1+\varepsilon})$.*

To prove Theorem 4 we need the following property that will be established in the final version of this paper.

▶ **Lemma 9.** *For all $\delta > 1/2$ there exists B such that*

$$\left|P_m^c(e^{\tau_1 m^{-\delta}}, e^{\tau_2 m^{-\delta}})\exp(-m^{-\delta}(\tau_1 \boldsymbol{E}[T_m^c] + \tau_2 \boldsymbol{E}[L_m^c]))\right| \leq B\sqrt{m}. \tag{15}$$

Now we proceed to prove Theorem 4. We apply Markov inequality for all θ and for all $x > 0$

$$P(|T_m^c - \mathbf{E}[T_m^c]| + |L_m^c - \mathbf{E}[L_m^c]| \geq 2xm^\delta) \leq P(|T_m^c - \mathbf{E}[T_m^c]| \geq xm^\delta \vee (|L_m^c - \mathbf{E}[L_m^c]| \geq xm^\delta)$$
$$\leq \left(P_m^c(e^\theta, 1)e^{-E[T_m^c]\theta} + P_m^c(e^{-\theta}, 1)e^{E[T_m^c])\theta}\right) e^{-x\theta m^\delta}$$
$$+ \left(P_m^c(1, e^\theta)e^{-E[L_m^c]\theta} + P_m^c(1, e^{-\theta})e^{E[L_m^c])\theta}\right) e^{-x\theta m^\delta}.$$

To complete the proof we will use (15) of Lemma 9. If we take $\tau_1 = \pm C$ and $\tau_2 = 0$ (and reverse) for some $C > 0$ such that $(\tau_1, \tau_2) \in \mathcal{U}(0)^2$, and $\theta = Cm^{-\delta'}$ for some $\delta' < \delta$, then we find $e^{\theta m^\delta} = e^{-Cm^\beta}$ with $\beta = \delta - \delta' > 0$, and

$$P(|T_m^c - \mathbf{E}[T_m^c]| + |L_m^c - \mathbf{E}[L_m^c]| \geq 2xm^\delta) \leq 4\sqrt{m}Be^{-xCm^\beta}$$

which prove (12) of Theorem 4. We can readjust by taking $0 < \beta' < \beta$ and the value of B to omit the factor $\sqrt{m}$.

4 Proof of Theorem 5 for LZ

We now consider the LZ78 algorithm over a single infinite sequence generated by a Markov source, that is, the n-LZ model and connect it to the n-DST model in which the path length is equal to n (over a variable number of independently inserted strings). In the m-LZ model there are exactly m LZ phrases, each being a block carved in the Markovian sequence. The blocks are *not* i.i.d Markovian sequences.

Let $\mathcal{P}_{m,n}^c$ be the probability that the length of the first m LZ phrases is exactly n (when the first symbol is c), leading to the n-LZ model. Notice that not every pair (n, m) is feasible in the LZ model since by adding another phrase the path length may "jump" by more than one. We are interested in finding an asymptotic estimate of $\mathcal{P}_{m,n}^c$. We start by introducing yet another model. Let $\mathbf{s}$ be a sequence of m symbols, namely $\mathbf{s} = (c_1, \ldots, c_m) \in \mathcal{A}^m$. For $c \in \mathcal{A}$ we now compute the probability $\mathcal{P}_{\mathbf{s},n}^c$ that an infinite Markovian sequence starting with symbol c when parsed by LZ algorithm satisfies the following two properties: (i) the first m blocks have tail symbols $c_i \in \mathbf{s}$ for $i \leq m$ so that c_i is the first symbol of block $i + 1$; (ii) the length of the first m LZ phrases is equal to n. If a string satisfies these two conditions, then we say it is $(\mathbf{s}, n)$ compatible and that it belongs to the $(\mathbf{s}, n)$-LZ model.

Given a string $\mathbf{s}$ of tail symbols we denote by $\mathbf{t}_c^a(\mathbf{s})$ (resp. $\mathbf{t}_c^b(\mathbf{s})$) the subsequence of $\mathbf{s}$ consisting of tail symbols of the LZ blocks starting with symbol a (resp. starting by symbol b). Now, it is easy to see that given the initial symbol c we can deduce the sequence of tails symbols and initial symbols of all phrases just by looking at the sequence $\mathbf{s}$, where the initial symbol of the next phrase is the tail symbol of the previous phrase. For example, if $\mathbf{s} = (a, b, a, b, b)$ and $c = a$ we have the following tail symbol and initial symbol sequence displayed in the following table:

block #	initial symbol	tail symbol
1	a	a
2	a	b
3	b	a
4	a	b
5	b	b

By taking the blocks (phrases) starting with $c = a$ we find $\mathbf{t}_a^a(\mathbf{s}) = (a, b, b)$ and the blocks starting with b yield $\mathbf{t}_a^b(\mathbf{s}) = (a, b)$.

Now we consider a sequence $\mathbf{t}$ of m symbols and introduce a new n-DST model which we call $(\mathbf{t}, n)$-DST model. We define by $P^c_{\mathbf{t},n}$ the probability that m i.i.d. (independent) Markovian sequences all starting with c satisfy the following two conditions (notice that we use roman P for this probability and calligraphic $\mathcal{P}$ for LZ model): (i) the tail symbol sequence follows the sequence $\mathbf{t}$; (ii) the external path length of the DST is exactly n. We will say that such m strings are $(\mathbf{t}, n)$-fit if they satisfy the above conditions and call it $(\mathbf{t}, n)$-DST model. We also define

$$P^c_{m,k,n} = \sum_{\mathbf{t}:\ |\mathbf{t}|=m, |\mathbf{t}|_a=k} P^c_{\mathbf{t},n} \tag{16}$$

with $|\mathbf{t}|$ being the length of sequence $\mathbf{t}$ and $|\mathbf{t}|_a$ being the number of symbols equal to a in it.

We finally establish the following fundamental lemma that connects the above two parameters which also connects the LZ parsing over a single Markovian sequence and the DST made of independent Markovian sequences, that is, $(\mathbf{s}, n)$-LZ model and $(\mathbf{t}, n)$-DST model where $\mathbf{t}$ is a function of $\mathbf{s}$.

▶ **Lemma 10.** *For any $\boldsymbol{s} \in \mathcal{A}^m$ we have*

$$\mathcal{P}^c_{\boldsymbol{s},n} = \sum_{n_a} P^a_{\boldsymbol{t}^a_c(\boldsymbol{s}),n_a} P^b_{\boldsymbol{t}^b_c(\boldsymbol{s}),n-n_a} \tag{17}$$

where n_c (equal either to n_a or n_b) is the path length in n_c-DST model with all strings starting with c, and $\boldsymbol{t}^a_c(\boldsymbol{s})$, $\boldsymbol{t}^b_c(\boldsymbol{s})$ are substrings of $\boldsymbol{s}$ as defined above.

Proof. In this conference paper, we give a proof using an example to ease the presentation. Let us consider $X = aabbababab\cdots$ which results in the following LZ blocks: $()(a)(ab)(b)(aba)(ba)(b\cdots)$. Or equivalently $X = \mathbf{aa}b\mathbf{ba}ba\mathbf{ba}b\mathbf{ab}\cdots$ where the initial block (phrase) symbols are displayed in bold. We notice that the first five blocks (excluding the initial empty block) accounts for a string of length 9. Thus the sequence X is $(\mathbf{s}, 9)$ compatible with $\mathbf{s} = (a, b, a, b, b)$. Given that X starts with symbol a we have $P(X) = P(\mathbf{a}|a)P(a\mathbf{a}|a)P(ab\mathbf{b}|a)P(b\mathbf{a}|b))P(aba\mathbf{b}|a)P(ba\mathbf{b}|b)$. Notice that we display in bold the tail symbol of each block (which is the initial symbol of the next block). We must incorporate $P(X)$ into $P^a_{\mathbf{s},9}$. In fact X should be viewed as the set of (infinite) strings having $aabbababab$ as the common prefix. We can rewrite $P(X)$ by regrouping the terms with respect to the initial symbol of each block as: $P(X) = [P(aa|a)P(abb|a)P(abab|a)] \times [P(ba|b)P(bab|b)]$. Observe that the sequence of strings $(aa, abb, abab)$ are the prefixes of a set of tuples of independent infinite strings that are all $(\mathbf{s}^a, 6)$ compatible with $\mathbf{s}^a = \mathbf{t}^a_a(\mathbf{s}) = (a, b, b)$ under the condition that the strings start with symbol a (the path length in the DST excludes the tail symbols, thus we must remove one from the length of each prefix). The probability of such event is exactly $P(aa|a)P(abb|a)P(abab|a)$ and must be incorporated in $P^a_{\mathbf{s}^a,6}$. Furthermore, these sequences are used to build one (left) part of the DST tree with independent Markov strings all starting with a. The same holds for the sequence of strings (ba, bab) which is $(\mathbf{s}^b, 3)$ compatible with $\mathbf{s}^b = \mathbf{t}^b_a(\mathbf{s}) = (a, b)$ and used to build the other part (right) of the DST tree. This leads to (17). ◀

The next crucial lemma connects n-LZ and n-DST models.

▶ **Lemma 11.** *The following holds*

$$\begin{aligned}\mathcal{P}^c_{m,n} \ \leq\ & \sum_{n_a}\sum_{k}\sum_{m_a} \big(P^a_{m_a,k,n_a} P^b_{m-m_a,m_a-k,n-n_a} \\ & + P^a_{m_a,k,n_a} P^b_{m-m_a,m_a-k-1,n-n_a} + P^a_{m_a,k,n_a} P^b_{m-m_a,m_a-k+1,n-n_a}\big)\end{aligned} \tag{18}$$

where n_a is the total path length of the first m_a phrases starting with an "a".

Proof. We naturally have $\mathcal{P}^c_{m,n} = \sum_{|\mathbf{s}|=m} \mathcal{P}^c_{\mathbf{s},n}$ where $|\mathbf{s}|$ is the length of the sequence $\mathbf{s}$. Similarly we have $P^c_{m,k,n} = \sum_{\mathbf{t},|\mathbf{t}|=m,|\mathbf{t}|_a=k} P^c_{\mathbf{t},n}$ with $|\mathbf{t}|_a$ is the number of symbols identical to a in $\mathbf{t}$. The rest follows from Lemma 10 but we need to take into account some boundary effects.

Let's look at it in more details. By (17) and above we find

$$\mathcal{P}^c_{m,n} = \sum_{|\mathbf{s}|=m} \sum_{n_a} P^a_{\mathbf{t}^a_c(\mathbf{s}),n_a} P^b_{\mathbf{t}^b_c(\mathbf{s}),n-n_a}.$$

We now partition $\mathcal{A}^m$into four sets $\mathcal{S}^c_0(m)$, $\mathcal{S}^c_1(m)$, $\mathcal{S}^c_2(m)$ and $\mathcal{S}^c_3(m)$:

- $\mathbf{s} \in \mathcal{S}^c_0(m)$: if neither of the initial symbol c or the final symbol of $\mathbf{s}$, namely c_m is identical to a. Thus the total number of tail symbols equal to a, namely $|\mathbf{s}|_a$ is equal to $|\mathbf{t}^a_c(\mathbf{s})|$.
- $\mathbf{s} \in \mathcal{S}^c_1(m)$: if both the final symbol and c are equal to a so that the total number of tail (and initial) symbols equal to a is $|\mathbf{t}^a_c(\mathbf{s})|$.
- $\mathbf{s} \in \mathcal{S}^c_2(m)$: if $c = a$ but $c_m \neq a$ so that the number of tail symbols equal to a is $|\mathbf{t}^a_c(\mathbf{s})| - 1$.
- $\mathbf{s} \in \mathcal{S}^c_3(m)$: if $c \neq a$ but the final symbol $c_m = a$. Thus the number of tail symbols equal to a is $|\mathbf{t}^a_c(\mathbf{s})| + 1$.

Regrouping we have

$$\mathcal{P}^c_{m,n} = \sum_{\mathbf{s}\in\mathcal{S}^c_0(m)\cup\mathcal{S}^c_1(m)} \mathcal{P}^c_{\mathbf{s},n} + \sum_{\mathbf{s}\in\mathcal{S}_2(m)} \mathcal{P}^c_{\mathbf{s},n} + \sum_{\mathbf{s}\in\mathcal{S}_3(m)} \mathcal{P}^c_{\mathbf{s},n}.$$

Now we have to deal with the right hand side of (18), that is, with the DST model. Let $\mathcal{T}_1(m)$ be the set of pairs of *arbitrary* sequences denoted as $(\mathbf{t}^a, \mathbf{t}^b)$ such that $|\mathbf{t}^a| + |\mathbf{t}^b| = m$ and $|\mathbf{t}^a|_a + |\mathbf{t}^b|_a = |\mathbf{t}^a|$. We notice that for $\mathbf{s} \in \mathcal{S}^c_1(m) \cup \mathcal{S}^c_2(m)$: $(\mathbf{t}^a_c(\mathbf{s}), \mathbf{t}^b_c(\mathbf{s})) \in \mathcal{T}_1(m)$, hence

$$\begin{aligned}\sum_{\mathbf{s}\in\mathcal{S}^c_0(m)\cup\mathcal{S}^c_1(m)} \mathcal{P}^c_{\mathbf{s},n} &= \sum_{n_a} \sum_{\mathbf{s}\in\mathcal{S}^c_0\cup\mathcal{S}^c_1(m)} P^a_{\mathbf{t}^a(\mathbf{s}),n_a} P^b_{\mathbf{t}^b(\mathbf{s}),n-n_a} \\ &\le \sum_{n_a} \sum_{(\mathbf{t}_a,\mathbf{t}_b)\in\mathcal{T}_1(m)} P^a_{\mathbf{t}^a,n_a} P^b_{\mathbf{t}^b,n-n_a}.\end{aligned}$$

Notice that we have an upper bound, since for some pair $(\mathbf{t}^a, \mathbf{t}^b)$ in $\mathcal{T}^c_1(m)$ there may not exist $\mathbf{s} \in \mathcal{S}^c_1(m) \cup \mathcal{S}^c_2(m)$ such that $\mathbf{t}^a = \mathbf{t}^a(\mathbf{s})$ and $\mathbf{t}^b = \mathbf{t}^b(\mathbf{s})$. For example, let $c = a$ and for $m = 4$ we set $\mathbf{t}^a = (a, b)$ and $\mathbf{t}^b = (b, a)$, so that $|\mathbf{t}^a|_a + |\mathbf{t}^b|_a = |\mathbf{t}^a|$ but it is impossible to find $\mathbf{s}$ such that $(\mathbf{t}^a_a(\mathbf{s}), \mathbf{t}^b(\mathbf{s})) = (\mathbf{t}^a, \mathbf{t}^b)$.

Thanks to (16) we have $\sum_{\mathbf{t}:\ |\mathbf{t}|=m,|\mathbf{t}|_a=k} P^c_{\mathbf{t},n} = P^c_{m,k,n}$ leading to

$$\sum_{(\mathbf{t}_a,\mathbf{t}_b)\in\mathcal{T}_1(m)} \sum_{n_a} P^a_{\mathbf{t}^a,n_a} P^b_{\mathbf{t}^b,n-n_a} = \sum_{m_a,k} P^a_{m_a,k,n_a} P^b_{m-m_a,m_a-k,n-n_a}.$$

This proves the first term in the right hand side of (18). To prove the other two terms we introduce $\mathcal{T}_2(m)$ as the set of pairs of sequence $(\mathbf{t}^a, \mathbf{t}^b)$ such that $|\mathbf{t}^a| + |\mathbf{t}^b| = m$ and $|\mathbf{t}^a|_a + |\mathbf{t}^b|_a = |\mathbf{t}^a| - 1$. In this case

$$\sum_{\mathbf{s}\in\mathcal{S}_2(m)} \mathcal{P}^c_{\mathbf{s},n} \le \sum_{n_a} \sum_{(\mathbf{t}^a,\mathbf{t}^b)\in\mathcal{T}_2(m)} P^a_{\mathbf{t}^a,n_a} P^b_{\mathbf{t}^b,n-n_a},$$

and the second term of (18) is proved. And finally with $\mathcal{T}_3(m)$ as the set of pairs of sequence $(\mathbf{t}^a, \mathbf{t}^b)$ such that $|\mathbf{t}^a| + |\mathbf{t}^b| = m$ and $|\mathbf{t}^a|_a + |\mathbf{t}^b|_a = |\mathbf{t}^a| + 1$, we establish the third term of (18). ◀

To finish the proof of Theorem 5 we now use the previous lemmas to upper bound $\mathcal{P}^c_{m,n}$. Let $\mathcal{P}^c_{m,n} \leq K^c_{m,n}(0) + K^c_{m,n}(1) + K^c_{m,n}(-1)$ with

$$K^c_{m,n}(i) = \sum_{m_a} \sum_{n_a} \sum_{k} P^a_{m_a,k,n_a} P^b_{m-m_a,m_a-k-i,n-n_a}.$$

To simplify our presentation we only study $K^c_{m,n}(0)$. First, we rewrite the bound in Theorem 4 for the DST model as follows: for $\delta > 1/2$ there exist B and C strictly positive such that

$$P^c_{m,k,n} \leq B \exp\left[-Cm^{-\delta}|k - \mathbf{E}[T^c_m]| - Cm^{-\delta}|n - \mathbf{E}[L^c_m]|\right].$$

Thus

$$\begin{aligned} K^c_{m,n}(0) \leq \textstyle\sum_{m_a+m_b=m} \sum_{k\leq m_a} \sum_{n_a+n_b=n} B^2 \exp\big[&-Cm_a^{-\delta}|k - \mathbf{E}[T^c_{m_a}] \\ &-Cm_a^{-\delta}|n_a - \mathbf{E}[L^a_{m_a}| - Cm_b^{-\delta}|m_a - k - \mathbf{E}[T^b_{m_b}] - Cm_b^{-\delta}|n_b - \mathbf{E}[L^b_{m_b}|\big]. \end{aligned}$$

From here we use $m_a, m_b \leq m$ to find

$$\begin{aligned} &Cm_a^{-\delta}|k - \mathbf{E}[T^c_{m_a}] + Cm_a^{-\delta}|n_a - \mathbf{E}[L^a_{m_a}| + Cm_b^{-\delta}|m_a - k - \mathbf{E}[T^b_{m_b}] + Cm_b^{-\delta}|n_b - \mathbf{E}[L^b_{m_b}| \geq \\ &Cm^{-\delta}|k - \mathbf{E}[T^c_{m_a}] + Cm^{-\delta}|n_a - \mathbf{E}[L^a_{m_a}| + Cm^{-\delta}|m_a - k - \mathbf{E}[T^b_{m_b}] + Cm^{-\delta}|n_b - \mathbf{E}[L^b_{m_b}| \\ &\geq Cm^{-\delta}|m_a - \mathbf{E}[T^a_{m_a}] - \mathbf{E}[T^b_{m_b}]| + Cm^{-\delta}|n - \mathbf{E}[L^a_{m_a}] - \mathbf{E}[L^b_{m_b}]|. \end{aligned}$$

Replacing the $\mathbf{E}[T^c_m]$ by $\tau_c(m)m$ and $\mathbf{E}[L^c_m]$ by $m \log m/h + m + m\mu_c(m)$ we arrive at

$$\begin{aligned} K^c_{m,n}(0) \leq B^2 m \sum_{m_a+m_b=m} &\exp\left(-Cm^{-\delta}|m_a - m_a\tau_a(m_a) - m_b\tau_b(m_b)|\right) \\ &\times \exp\left(-Cm^{-\delta}|n - m\log m/h + m(H(m_a/m)/h - 1) - m_a\mu_a(m_a) - m_b\mu_b(m_a)|\right). \end{aligned}$$

Without changing the order of magnitude we further can replace $\tau_c(m)$ by $\tau(m)$ and $\mu_c(m)$ by $\mu(m)$.

We now focus only on the aperiodic case and set $\tau(m) = \bar{\tau}m$ and $\mu(m) = \bar{\mu}m$. (We know that even in this case for small values of m, the $\mu(m)$ and $\tau(m)$ are not exactly linear in m, but we handle it later.) Thus our term $K^c_{m,n}(0)$ is bounded by

$$B^2 m \sum_{m_a \leq m} \exp[-Cm^{-\delta}|m_a - \bar{\tau}m|] \exp[-Cm^{-\delta}|n - m\log m/h - \bar{\mu}m + m(H(m_a/m)/h - 1)|].$$

If we take any $\delta' > \delta$ we find

$$\begin{aligned} K^c_{m,n}(0) \leq B^2 m \sum_{m_a \leq m} &\exp[-Cm^{-\delta}|m_a - \bar{\tau}m|] \\ &\times \exp[-Cm^{-\delta'}|n - m\log m/h - \bar{\mu}m + m(H(m_a/m)/h - 1)|]. \end{aligned}$$

We observe that $\exp[-Cm^{-\delta}|m_a - \bar{\tau}m|]$ attains its maximum at $m_a = m^* = \bar{\tau}m$. Thus

$$\begin{aligned} K^c_{m,n}(0) \leq B^2 \sum_{m_a \leq m^*} &e^{Cm^{-\delta}(m-m^*)} \times \exp[-Cm^{-\delta'}|n - m\log m/h - \bar{\mu}m + m(H(m_a/m)/h - 1)|]] \\ +B^2 \sum_{m_a \geq m^*} &e^{Cm^{-\delta}(m^*-m)} \times \exp[-Cm^{-\delta'}|n - m\log m/h - \bar{\mu}m + m(H(m_a/m)/h - 1)|]]. \end{aligned}$$

Notice that the terms $e^{Cm^{-\delta}(m-m^*)}$ and $e^{Cm^{-\delta}(m^*-m)}$ form a geometrically decreasing series with rate $e^{-Cm^{-\delta}}$. Since $|mH((m_a+1)/m) - mH(m_a/m)| \le \log m$, the term

$$\exp[-Cm^{-\delta'}|n - m\log n/h - \bar{\mu}m + m(H(m_a/m)/h - 1)|]]$$

is at most geometrically increasing with a rate $e^{m^{-\delta'}\log m/h}$ which is smaller than $e^{Cm^{-\delta}}$. Therefore, the whole series has its maximum at $m_a = m^*$ and

$$K^c_{m,n}(0) \le 2B^2 \sum_{k=0^\infty} e^{-Ck(m^{-\delta} - \log m/hm^{-\delta'})}$$

$$\times \exp[-Cm^{-\delta'}|n - m\log n/h - \bar{\mu}m + m(H(m^*/m)/h - 1)|]]$$

$$= \frac{2B^2}{1 - e^{-(m^{-\delta} - \log m/hm^{-\delta'})C}}$$

$$\times \exp[-Cm^{-\delta'}|n - m\log m/h - \bar{\mu}m + m(H(m^*/m)/h - 1)|]]$$

$$= O(2B^2m^\delta)\exp[-Cm^{-\delta'}|n - m\log m/h - \bar{\mu}m + m(H(\bar{\tau})/h - 1)|]].$$

Including all contributions, the final estimate for some $B' > 0$ is

$$\mathcal{P}^c_{m,n} \le B'm^{1+\delta}\exp[-Cm^{-\delta}|n - m\log m - \bar{\mu}m + m(H(\bar{\tau})/h - 1)|].$$

This gives the large deviation estimate and $\mathbf{E}[\mathcal{L}^c_{m,n}] = m\log m/h + \bar{\mu}m - m(H(\bar{\tau})/h - 1) + O(m^\delta)$ by Fact 1. We recognize in $H(\bar{\tau})$ the entropy of the tail symbol.

In fact the quantities $\tau(m)$ and $\mu(m)$ are not exactly $\bar{\tau}m$ and $m\bar{\mu}$. To handle it we observe that due to their slowly varying properties, the function $\exp(-Cm^{-\delta}|m_a - \tau(m_a)m_a - \tau(m - m_b)(m - m_a)|$ attains the maximum for m^* such that

$$m^* = -\tau_a(m^*)m^* - \tau_b(m^*)(m - m^*).$$

Indeed the function $m_a - \mathbf{E}[T^a_{m_a}] - \mathbf{E}[T^b_{m_b}]$ is strictly increasing thus this value is unique. Then again $\mathbf{E}[\mathcal{L}^c_m] = m\log m/h + m^*\mu(m^*) + (m - m^*)\mu(m - m^*) - m(H(m^*/m)/h - 1)$, and therefore $\mathbf{E}[L^c_m] + mH(m^*/m) + o(m)$. The latter is equal to $\mathbf{E}[L^c_m] + mH(\bar{\tau}) + o(m)$ in the aperiodic case. To complete the proof of Theorem 5 we just use Fact 1 applied to $\mathcal{L}_m$.

5 Conclusions

In this paper we analyze the Lempel-Ziv'78 algorithm for binary Markov sources, a problem left open since the algorithm inception. To handle the strong dependency between Markov phrases, we introduce and precisely analyze the so called tail symbol which is the first symbol of the next phrase in the LZ78 parsing. We focus here on the large deviations for the number of phrases in the LZ78 and also give a precise asymptotic expression for the redundancy which is the excess of LZ78 code over the entropy of the source. In future work we plan to extend our analysis to non-binary Markov sources and present some bounds on the central limit theorem. Furthermore, we shall study LZ78 for Markov sources of higher order, however, this will require a new approach to the tail symbols which may span over consecutive phrases.

References

1 D. Aldous and P. Shields. A diffusion limit for a class of random-growing binary trees. *Probab. Th. Rel. Fields*, 79:509–542, 1988.

2 J. Fayolle and M. Ward. Analysis of the average depth in a suffix tree under a markov model. In *2005 International Conference on Analysis of Algorithms*, pages 95–104, 2005.

3 P. Jacquet and W. Szpankowski. Asymptotic behavior of the lempel-ziv parsing scheme and digital search trees. *Theoretical Computer Science*, 144:161–197, 1995.

4 P. Jacquet and W. Szpankowski. Anaylytical depoissonization and its applications. *Theoretical Computer Science*, 201:1–62, 1998.

5 P. Jacquet and W. Szpankowski. On the limiting distribution of lempel ziv'78 redundancy for memoryless sources. *IEEE Trans. Information Theory*, 60:6917–6930, 2014.

6 P. Jacquet and W. Szpankowski. *Analytic Pattern Matching: From DNA to Twitter*. Cambridge University Press, Cambridge, 2015.

7 P. Jacquet, W. Szpankowski, and J. Tang. Average profile of hte lempel-ziv parsing scheme for a markovian source. *Algorithmica*, 31:318–360, 2001.

8 D. E. Knuth. *The Art of Computer Programming Sorting and Searching*. Addison-Wesley Reading, 1998.

9 K. Leckey, R. Neininger, and W. Szpankowski. Towards more realistic probabilistic models for data structures: The external path length in tries under the markov model. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 877–886, 2013.

10 G. Louchard and W. Szpankowski. Average profile and limiting distribution for a phrase size in the lempel-ziv parsing algorithm. *IEEE Trans. Information Theory*, 41:478–488, 1995.

11 N. Merhav. Universal coding with minimum probability of codeword length overflow. *IEEE Trans. Information Theory*, 37:556–563, 1991.

12 N. Merhav and J. Ziv. On the amount of statistical side information required for lossy data compression. *IEEE Trans. Information Theory*, 43:1112–1121, 1997.

13 R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14:378–418, 2004.

14 S. Savari. Redundancy of the lempel-ziv incremental parsing rule. *IEEE Trans. Information Theory*, 43:9–21, 1997.

15 W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley New York, New York, 2001.

16 J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Information Theory*, 24:530–536, 1978.

A Proof of Theorem 3(i): Mean

We first analyze asymptotically $\mathbf{X}(z) = (X_a(z), X_b(z))$ that satisfies the system of differential-functional equations (6). We solve this system, and then apply Mellin transform and depoissonization to prove Theorem 3(i).

Since for all integer m, we have $T_m^c \leq m$, we notice that the function $X_c(z)$ is $O(z)$ both when $z \to \infty$ and when $z \to 0$. Thus the function $\mathbf{X}(z)$ has no Mellin transform defined as $X_c(s) = \int_0^\infty X_c(z) z^{s-1} dz$ (see [15] for more on the Mellin transform). To correct this we introduce $\tilde{X}_c(z) = X_c(z) - G_c(z)$ with $G_c(z) = (\mathbf{E}[T_1^c]z + \mathbf{E}[T_2^c]z^2/2)e^{-z}$ which is $O(z^3)$ when $z \to 0$, where $\mathbf{E}[T_1^c]$ and $\mathbf{E}[T_2^c]$ are defined in (7).

The Mellin transform $X_c^*(s)$ of $\tilde{X}_c(z)$ on the strip $\Re(s) \in]-3, -1[$ exists. The Mellin transform of $\partial_z \tilde{X}_c(z)$ exists too on the strip $\Re(s) \in]-2, 0[$. Thus the two Mellin transforms coexist on the strip $\Re(s) \in]-2, -1[$ and satisfies [15]

$$- (s-1)(X_c^*(s-1) + G_c^*(s)) + X_c^*(s) + G_c^*(s)$$
$$= P(a|c)^{-s}(X_a^*(s) + G_a^*(s)) + P(b|c)^{-s}(X_b^*(s) + G_b^*(s))$$

where $G_c^*(s)$ for $c \in \mathcal{A}$ is the Mellin transform of $G_c(z)$ and has the explicit expression $\mathbf{E}[T_1^c]\Gamma(1+s) + \mathbf{E}[T_2^c]\Gamma(s+2)/2$. This expression is here for completeness.

An alternative but convenient way to see this equations is to consider the vector $\mathbf{X}^*(s)$ made of the quantities $X_c^*(s)$, $c \in \mathcal{A}$ which is also the Mellin transform of the vector $\tilde{\mathbf{X}}(z)$ made of the coefficients $\tilde{X}_c(z)$. This yields the linear equation

$$-(s-1)(\mathbf{X}^*(s-1) + \mathbf{G}^*(s-1)) + \mathbf{X}^*(s) + \mathbf{G}^*(s) =$$

$$= \mathbf{P}(s)(\mathbf{X}^*(s) + \mathbf{G}^*(s))$$

where $\mathbf{G}^*(s)$ is the vector of the $G_c^*(s)$. It can be rewritten in

$$(s-1)(\mathbf{X}^*(s-1) + \mathbf{G}^*(s-1)) = (\mathbf{I} - \mathbf{P}(s))(\mathbf{X}^*(s) + \mathbf{G}^*(s)).$$

This kind of equation has been studied in [7] where we introduce a new function $\mathbf{x}(s)$

$$\mathbf{X}^*(s) + \mathbf{G}^*(s) = \Gamma(s)\mathbf{x}(s).$$

Thus the equation becomes $\mathbf{x}(s-1) = (\mathbf{I} - \mathbf{P}(s))\mathbf{x}(s)$, which leads to $\mathbf{x}(s) = \prod_{i\geq 0}(\mathbf{I} - \mathbf{P}(s-i))^{-1}\mathbf{K}$ where $\mathbf{K}$ is a constant vector. Notice that the matrices very likely don't commute thus the product order is specified from the left to right. Indeed we have

$$\mathbf{K} = \left(\prod_{j\geq 2}(\mathbf{I} - \mathbf{P}(-j))^{-1}\right)^{-1}\mathbf{x}(-2) = \prod_{j=-\infty}^{j=2}(\mathbf{I} - \mathbf{P}(j))\mathbf{x}(-2). \tag{A.1}$$

To handle it we need an explicit formula for $\mathbf{x}(-2)$. The following lemma from [7] is useful in this regard. We provide a proof for completeness.

▶ **Lemma 12.** *Let $\{f_n\}_{n=0}^{\infty}$ be a sequence of real numbers having the Poisson transform*

$$\widetilde{F}(z) = \sum_{n=0}^{\infty} \tilde{f}_n \frac{z^n}{n!} e^{-z} := \sum_{n=0}^{\infty} f_n \frac{z^n}{n!}, \tag{A.2}$$

which is an entire function. Furthermore, let its Mellin transform $F(s)$ have the following factorization

$$F(s) = \mathcal{M}[\widetilde{F}(z); s] = \Gamma(s)\gamma(s).$$

Assume that $F(s)$ exists for $\Re(s) \in (-2, -1)$, and that $\gamma(s)$ is analytic for $\Re(s) \in (-\infty, -1)$. Then

$$\gamma(-n) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k \tilde{f}_k = (-1)^n f_n, \quad \text{for} \quad n \geq 2. \tag{A.3}$$

Proof. Notice that f_n and $\tilde{f}_n$ are related by [15]

$$\tilde{f}_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} f_k \ , \quad n \geq 0 \ .$$

Define for some fixed $M \geq 2$, the function $\widetilde{F}_M(z) = \sum_{n=0}^{M-1} f_n \frac{z^n}{n!}$. Due to our assumptions, we can continue $F(s)$ analytically to the whole complex plane except $s = -2, -3, \ldots$. In particular, for $\Re(s) \in (-M, -M+1)$ we have $F(s) = \mathcal{M}[\widetilde{F}(z) - \widetilde{F}_M(z); s]$. As $s \to -M$, due to the factorization $F(s) = \Gamma(s)\gamma(s)$, we have

$$F(s) = \frac{1}{s+M} \frac{(-1)^M}{M!} \gamma(-M) + O(1) \; ;$$

thus by the inverse Mellin transform, we have

$$\widetilde{F}(z) - \widetilde{F}_M(z) = \frac{(-1)^M}{M!} \gamma(-M) z^M + O(z^{M+1}) \quad \text{as} \quad z \to 0 \; . \tag{A.4}$$

But

$$\widetilde{F}(z) - \widetilde{F}_M(z) = \sum_{i=M}^{\infty} f_n \frac{z^n}{n!} = f_M \frac{z^M}{M!} + O(z^{M+1}) \; . \tag{A.5}$$

Comparing (A.4) and (A.5) shows that $\gamma(-M) = (-1)^M f_M = \sum_{k=0}^{M} \binom{M}{k} (-1)^k \tilde{f}_k$. ◀

Now we can compute $\mathbf{x}(-2)$ using above and (7) leading to

$$\mathbf{x}(-2) = \begin{bmatrix} T_2^a - 2P(a|a) \\ T_2^b - 2P(a|b) \end{bmatrix}. \tag{A.6}$$

In another notation $\mathbf{x}(-2) = (\mathbf{P}^2 - \mathbf{P})\mathbf{e}_a$, where $\mathbf{e}_a$ is the vector made of a single 1 at a position and zero otherwise.

Next, we notice that the vector

$$\Gamma(s) \prod_{i \geq 0} (\mathbf{I} - \mathbf{P}(s-i))^{-1} \prod_{j=-\infty}^{j=-2} (\mathbf{I} - \mathbf{P}(j)) \mathbf{x}(-2)$$

may have a double pole on $s = -1$ since $\Gamma(s)$ has a pole and also $(\mathbf{I} - \mathbf{P}(s))^{-1}$ since $\mathbf{I} - \mathbf{P}(-1) = \mathbf{I} - \mathbf{P}$ is singular. But in fact the pole multiplicity is reduced by one, as prove below. Let us also define

$$\mathbf{Q}(s) = \prod_{i \geq 1} (\mathbf{I} - \mathbf{P}(s-i))^{-1} \prod_{j=-\infty}^{j=-2} (\mathbf{I} - \mathbf{P}(j)).$$

Then $\mathbf{x}(s) = (\mathbf{I} - \mathbf{P}(s))^{-1} \mathbf{Q}(s) \mathbf{x}(-2)$.

We notice that when $s \to -1$, then $\mathbf{Q}(s) = \mathbf{I} + (s+1)\mathbf{Q}'(-1) + O((s+1)^2)$. Furthermore let $\lambda(s)$ be the main eigenvalue of matrix $\mathbf{P}(s)$ and $\mathbf{1}(s)$ and $\boldsymbol{\pi}(s)$ be respectively the right and left main eigenvectors. We have $\lambda(-1) = 1$, $\mathbf{1}(-1) = \mathbf{1}$ is all made of one's, and $\boldsymbol{\pi}(-1)$ is the stationary distribution of the Markov source.

From the matrix spectral representation [15] we have

$$\mathbf{P}(s) = \lambda(s)\mathbf{1}(s) \otimes \boldsymbol{\pi}(s) + \mathbf{R}(s) = \lambda(s)\boldsymbol{\Pi}(s) + \mathbf{R}(s) \tag{A.7}$$

where $\mathbf{R}(s)$ is the automorphism of the eigenplan orthogonal to the main eigenvector and $\boldsymbol{\Pi}(s) = \mathbf{1}(s) \otimes \boldsymbol{\pi}(s)$ where $\otimes$ is the tensor product. Note that $\boldsymbol{\Pi} \cdot \mathbf{P} = \mathbf{P} \cdot \boldsymbol{\Pi} = \boldsymbol{\Pi}$. Then

$$\begin{aligned} (\mathbf{I} - \mathbf{P}(s))^{-1} \quad = \quad & \frac{1}{1 - \lambda(s)} \mathbf{1}(-s) \otimes \boldsymbol{\pi}(s) \\ & - \frac{1}{\lambda'(-1)} \left(\mathbf{1}'(-1) \otimes \boldsymbol{\pi}(-1) + \mathbf{1} \otimes \boldsymbol{\pi}'(-1) \right) + \mathbf{R}(-1)^{-1} + O(s+1). \end{aligned}$$

Finally

$$(\mathbf{I}-\mathbf{P}(s))^{-1}\mathbf{Q}(s)\mathbf{x}(-2) = \frac{\mathbf{1}\otimes\boldsymbol{\pi}(s)(\mathbf{I}-\mathbf{P})\mathbf{e}_a}{1-\lambda(s)} - \frac{1}{\lambda'(-1)}\left(\mathbf{1}'(-1)\otimes\boldsymbol{\pi}+\mathbf{1}\otimes\boldsymbol{\pi}'(-1)\right) + \mathbf{R}^{-1}(-1) + \frac{(s+1)}{1-\lambda(s)}\mathbf{1}\otimes\mathbf{Q}'(-1)+O(s+1).$$

Since

$$\frac{s+1}{1-\lambda(s)} \to -\frac{1}{\lambda'(-1)}$$

when $s\to -1$, and $\boldsymbol{\Pi}\mathbf{P}(\mathbf{I}-\mathbf{P})\mathbf{e}_a = (\boldsymbol{\Pi}-\boldsymbol{\Pi})\mathbf{e}_a = 0$. Also

$$\mathbf{R}^{-1}(-1)(\mathbf{I}-\mathbf{P})\mathbf{P}\mathbf{e}_a = \mathbf{P}\mathbf{e}_a - \langle\boldsymbol{\pi}\mathbf{P}\mathbf{e}_a\rangle\mathbf{1} = \mathbf{P}\mathbf{e}_a - \langle\boldsymbol{\pi}\mathbf{e}_a\rangle\mathbf{1}. \tag{A.8}$$

We finally have

$$\lim_{s\to -1}\mathbf{x}(s) = \mathbf{P}\mathbf{e}_a - \pi_a\mathbf{1} - \frac{1}{\lambda'(-1)}\mathbf{1}\langle\left(\boldsymbol{\pi}'(-1)+\boldsymbol{\pi}\mathbf{Q}'(-1)\right)(\mathbf{I}-\mathbf{P})\mathbf{P}\mathbf{e}_a\rangle, \tag{A.9}$$

where π_a is the coefficient of the stationary distribution $\boldsymbol{\pi}$ at symbol a.

Now we are in position to establish asymptotics of $X_c(z)$ for large z and through depoissonization asymptotics of $\mathbf{E}[T_m^c]$. The inverse Mellin transform is

$$\tilde{X}_c(z) = \frac{1}{2i\pi}\int_{x-i\infty}^{x+i\infty} X_c^*(s)z^{-s}ds \tag{A.10}$$

valid for all $x\in]-2,-1[$. Remembering that $T_c(z) = \tilde{X}_c(z) + P(a|c)z$ we have indeed

$$\tilde{\mathbf{X}}(z) = \frac{1}{2i\pi}\int_{x-i\infty}^{x+i\infty}\Gamma(s)\mathbf{x}(s)z^{-s}ds - \frac{1}{2i\pi}\int_{x-i\infty}^{x+i\infty}\mathbf{G}^*(s)z^{-s}ds. \tag{A.11}$$

We know that $\mathbf{T}(z)-\tilde{\mathbf{X}}(z)$ is decaying exponentially fast when $z\to\infty$.

Moving the line of integration toward the right, we meet a single pole at $s=-1$ of $\mathbf{G}^*(s)z^{-z}$ and its residues is $-z\mathbf{P}\mathbf{e}_a$. Then

$$\frac{1}{2i\pi}\int_{x-i\infty}^{x+i\inf ty}\mathbf{G}^*(s)z^{-s}ds = -\mathbf{P}\mathbf{e}_a + O(z^{-M})$$

for all $M>0$.

The value -1 is also a simple pole for $z^{-s}\Gamma(s)\mathbf{x}(s)$. We know that its residue is

$$-z\left(\mathbf{P}\mathbf{e}_a - \pi_a\mathbf{1} - \frac{1}{\lambda'(-1)}\mathbf{1}\langle\left(\boldsymbol{\pi}'(-1)+\boldsymbol{\pi}\mathbf{Q}'(-1)\right)(\mathbf{I}-\mathbf{P})\mathbf{P}\mathbf{e}_a\rangle\right). \tag{A.12}$$

Therefore we have

$$\mathbf{X}(z) = z\left(\pi_a + \frac{1}{\lambda'(-1)}\mathbf{1}\langle\left(\boldsymbol{\pi}'(-1)+\boldsymbol{\pi}\mathbf{Q}'(-1)\right)(\mathbf{I}-\mathbf{P})\mathbf{P}\mathbf{e}_a\rangle\right)\mathbf{1} + o(z). \tag{A.13}$$

For irrational case, we know that $s=-1$ is the only pole on the line $\Re(s)=-1$, leading to the error term $o(z)$ coming from other poles of $(\mathbf{I}-\mathbf{P}(s))^{-1}$ which may occur on the right half plan of $s=-1$.

But in the rational case, there is the possibility of other poles regularly spaced on the axis $\Re(s)=-1$ with some specific matrices $\mathbf{P}$ detailed in [7] where the coefficients α_{abc} are introduced. In these very specific cases (the uniform probability distribution on $\mathcal{A}$ is one of them) the $o(z)$ term should be replaced by a term $zQ_c(\log z)+O(z^{1-\epsilon})$, where Q_c is a periodic vector of very small amplitude and mean zero, and $\epsilon>0$ depends on the matrix $\mathbf{P}$. This proves Theorem 3(i).

B Proof of Theorem 3(ii): Variance

We now analyze asymptotically $\mathbf{V}(z) = (V_a(z), V_b(z))$ that satisfies the system of differential-functional equations (8). In order to apply depoissonization, for $\theta \in [0, \pi/2]$ we define $\mathcal{C}(\theta)$ as the complex cone containing the complex number z such that $|\arg(z)| \leq \theta$ on increasing domains [15, 5]

$$\mathcal{C}_k(\theta) = \{z, z \in \mathcal{C}(\theta) \& |z| \leq \rho^k\}$$

with $\rho = \min_c\{\frac{1}{P(a|c)}, \frac{1}{P(b|c)}\}$.

Our first goal is to prove that $V_c(z) = O(z)$. We shall use use the increasing domain approach [15] applied to (8) following the footsteps of the proof of Lemma 7A of [3]. From Fact 1 of [3] we conclude that

$$V_c(z) = V_c(\rho z)e^{-z(1-\rho)} + e^{-z}\int_{\rho z}^{z} e^x \left(V_a(P(a|c)x) + V_b(P(b|c)x) + g(x)\right) dx \tag{B.14}$$

where $g(z) = P(a|c) - P^2(a|c) + [X_z^c(z)]^2 = O(1)$. Indeed, it follows from Fact 1 of [3] that the differential equation like

$$f'(z) = b(z) - a(z)f(z) \tag{B.15}$$

satisfies

$$f(z) = f(z_0)e^{A(z_0)-A(z)} + \int_{z_0}^{z} b(x)e^{A(x)-A(z)}dx$$

where $A(z) = \int a(z)$ is the primitive function of $a(z)$. Setting in (B.15) $f(z) = V_c(z)$, $b(z) = V_a(P(a|c)z) + V_b(P(b|c)z) + g(z)$ and $a(z) = 1$ we obtain (B.14).

Now we apply induction over the increasing domains. In short, we assume that for $z \in \mathcal{C}_k(\theta)$ we have $|V_c(z)| \leq B_k|z|$ for some B_k. Using the induction of the increasing domains we prove, as in the Appendix of [3] that B_k are bounded. This completes the proof, after applying the depoissonization lemma of [4].

In order to find a precise estimate of the asymptotic development of $\mathbf{V}(z)$ we denote $\mathbf{V}^*(s)$ the Mellin transform of $\mathbf{V}(z)$. From (8) we arrive at

$$-(s-1)\mathbf{V}^*(s-1) + \mathbf{V}^*(s) = \mathbf{P}(s)\mathbf{V}^*(s) + \mathbf{g}^*(s),$$

where $\mathbf{g}^*(s)$ is the Mellin transform of the vector made of the coefficients $(\partial_z X_c(z))^2$. Let $\mathbf{V}^*(s) = \Gamma(s)\mathbf{B}(s)$ and $\mathbf{g}^*(s) = \Gamma(s)\mathbf{G}(s)$. Then

$$\mathbf{B}(s) = (\mathbf{I} - \mathbf{P}(s))^{-1}\left(\mathbf{B}(s-1) + \mathbf{G}(s)\right).$$

The quantity $(\mathbf{I} - \mathbf{P}(s))^{-1}$ has a pole at $s = -1$. Together with $\Gamma(s)$ it would give a double pole at $s = -1$ which is not possible, as proved above. Indeed, notice that the coefficient at the double pole at $s = 1$ is $\mathbf{\Pi}(\mathbf{B}(-2) + \mathbf{G}(-1)$. But $\mathbf{G}(-1)$ is the the coefficient at z of $\mathbf{g}(z)$ and $\mathbf{B}(-2)$ is the coefficient at z^2 of $\mathbf{V}(z)$, as already proved in Lemma 12. Then we easily see that $\mathbf{B}(-2) + \mathbf{G}(-1) = \mathbf{P}^2\mathbf{e}_a - \mathbf{P}\mathbf{e}_a$, and consequently the coefficient at the double pole at $s = 1-$ is equal to $\mathbf{\Pi}(\mathbf{P}^2\mathbf{e}_a - \mathbf{P}\mathbf{e}_a) = (\mathbf{\Pi} - \mathbf{\Pi})\mathbf{e}_a = 0$, as desired.

Therefore, the contribution of pole $s = -1$ to the asymptotic of $\mathbf{V}(z)$ is $\mathbf{B}(-1)$ becomes

$$\begin{aligned}\mathbf{B}(-1) &= \frac{1}{\lambda'(-1)}\left(\langle\boldsymbol{\pi}'(-1)(\mathbf{B}(-2) + \mathbf{G}(-1))\rangle + \langle\boldsymbol{\pi}(\mathbf{B}'(-2) + \mathbf{G}'(-1))\rangle\right)\mathbf{1} \\ &\quad + (\mathbf{I} - \mathbf{R}(-1))^{-1}(\mathbf{B}(-2) + \mathbf{G}(-1)).\end{aligned}$$

Notice also that $(\mathbf{I}-\mathbf{R}(-1))^{-1}(\mathbf{P}^2\mathbf{e}_a-\mathbf{P}\mathbf{e}_a)=\langle\boldsymbol{\pi}\mathbf{P}\mathbf{e}_a\rangle\mathbf{1}-\mathbf{P}\mathbf{e}_a=\langle\boldsymbol{\pi}\mathbf{e}_a\rangle\mathbf{1}-\mathbf{P}\mathbf{e}_a$.

The real issue here is how to compute $\mathbf{B}'(-2)$ and $\mathbf{G}'(-1)$, which we address next.

▶ **Lemma 13.** *Let a function* $g(z)=\sum_{n\geq 1}\frac{a_n}{n!}z^n$ *and* $f(z)=g(z)e^{-z}=\sum_{n\geq 1}\frac{b_n}{n!}z^n$. *Let also* $g_k(z)=\sum_{n\leq k}\frac{a_n}{n!}z^n$ *and* $f_k(z)=f(z)-g_k(z)e^{-z}$ *with* $f_k^*(s)$ *being its Mellin transform defined for* $-k-1<\Re(s)<0$. *Then*

$$\begin{aligned}\lim_{s\to -k}\left(\frac{f^*(s)}{\Gamma(s)}\right)' &= f_k^*(-k)\left(\frac{1}{\Gamma(s)}\right)'_{s=-k}+\sum_{n\leq k}\frac{a_n}{n!}\left(s^{\langle n\rangle}\right)'_{s=-k}\\ &= f_k^*(-k)(-1)^{n-1}n!+\sum_{n\leq k}\frac{a_n}{n!}\left(s^{\langle n\rangle}\right)'_{s=-k}\end{aligned}$$

where $s^{\langle n\rangle}=\frac{\Gamma(s+n)}{\Gamma(s)}=(s+n-1)\times\cdots\times s$.

Proof. We start with a simple identity

$$\frac{f^*(s)-f_k^*(s)}{\Gamma(s)}=\sum_{n\leq k}\frac{a_n}{n!}s^{\langle n\rangle}$$

which is easy to derive. But the Mellin transform of $f_k(z)$ and $f_k^*(s)$ are defined for $-k-1<\Re(s)<0$. The derivative of $f_k^*(s)/\Gamma(s)$ at $s=-k$ is equal to $f_k^*(-k)\left(\Gamma^{-1}(s)\right)'_{s=-k}$ since $\Gamma^{-1}(-k)=0$. Finally we notice that [15]

$$\lim_{s\to -k}\left(\frac{1}{\Gamma(s)}\right)'=\lim_{s\to -k}\frac{\Psi(s)}{\Gamma(s)}=\lim_{s\to -k}\frac{(s+n)\Psi(s)}{(s+n)\Gamma(s)}=(-1)^{n-1}n!$$

where $\Psi(s)$ is the psi function. ◀

In absence of specific properties on $f_k(z)$ there is no other way than numerical computation to get an estimate of $f_k^*(-k)$. Finally, we can present a precise asymptotic expression for the variance.

▶ **Theorem 14.** *We have* $\boldsymbol{V}(z)=\bar{\omega}_a\boldsymbol{1}z+o(z)$ *in the aperiodic case, and in the periodic case* $\boldsymbol{V}(z)=\bar{\omega}_a\boldsymbol{1}z+Q_2(\log z)z+O(z^{1-\epsilon})$ *for some* $\epsilon>0$ *and* $Q_2(.)$ *being a periodic function of small amplitude and mean zero, where*

$$\bar{\omega}_a=\frac{1}{\lambda'(-1)}\left(\langle\boldsymbol{\pi}'(-1)((\boldsymbol{P}-\boldsymbol{I})\boldsymbol{P}\boldsymbol{e}_a\rangle+\langle\boldsymbol{\pi}(\boldsymbol{B}'(-2)+\boldsymbol{G}'(-1))\rangle\right)+\langle\boldsymbol{\pi}\boldsymbol{e}_a\rangle. \tag{B.16}$$

Notice that $\omega=B(-1)+\boldsymbol{P}\boldsymbol{e}_a$.